Themenblock Datenbeschaffung und -aufbereitung: 04_03Aufbereitung_Pseudonymisierung

Pseudonymisierung und Anonymisierung

Erarbeitet von
Dr. Ann-Kathrin Selker

Die Inhalte dieses Videos stellen keinesfalls eine Rechtsberatung in irgendeiner Form dar oder rechtliche Leitlinien. Ziels dieses Videos ist es, ein Problembewusstsein im Umgang mit KI zu schaffen und für rechtliche Fragen in diesem Kontext zu sensibilisieren. Vor dem Einsatz von KI-Systemen im Rahmen deines Projekts oder deiner Arbeit wende dich an die jeweiligen Fachstellen deiner Universität oder deines Unternehmens, um die rechtlichen Rahmenbedingungen für den Einsatz von KI zu besprechen.

Lernziele

- Du kannst die Begriffe der Pseudonymisierung und Anonymisierung definieren
- Du kannst die Unterschiede zwischen Pseudonymisierung und Anonymisierung nennen
- Du kannst die Bedeutung von Pseudonymisierung und Anonymisierung beim Maschinellen Lernen erläutern

Projekt ai4all „KI für Alle"
HeiCAD –Heine Center for Artificial Intelligence and Data Science

Inhalt

Einstieg

Unsere gesammelten Trainingsdaten enthalten oft personenbezogene Daten, die nach der Datenschutzgrundverordnung geschützt sind. Was für Möglichkeiten gibt es diese zu schützen? Und sind diese Daten dann wirklich geschützt?

Pseudonymisierung

Um personenbezogene Daten zu schützen, können sie z. B. pseudonymisiert werden. Unter Pseudonymisierung verstehen wir das Ersetzen von personenbezogenen Daten durch „fiktive Daten", sogenannte Pseudonyme. Das betrifft unter anderem Namen, die häufig durch IDs ersetzt werden, schließt aber auch alle anderen personenbezogenen Daten mit ein. Dabei wird in einer Art Zuordnungsliste gespeichert, welche Realdaten durch welche Pseudonyme ersetzt wurden, sodass eine nachträgliche Zuordnung der Datenbeobachtungen immer noch möglich ist. Nach der DSGVO muss diese Zuordnungsliste sicher getrennt von den anderen Daten aufbewahrt und mit technischen sowie organisatorischen Maßnahmen gegen eine erneute Zuweisung gesichert werden.

Quelle [1]

Aber pseudonymisierte Daten können trotz Namensentfernung auch ohne die interne Zuordnungsliste identifizierbar sein. So gab Netflix im Jahr 2006 im Zuge eines Wettbewerbs einen Datensatz mit Rankings von Filmen durch Nutzer*innen heraus, bei denen die Benutzer*innennamen und Filmtitel durch IDs ersetzt wurden und zusätzlich die Daten noch leicht verändert wurden. Kurz darauf wurde aufgezeigt, dass trotzdem einzelne Nutzer*innen den veröffentlichten Datenpunkten zugeordnet werden konnten.

Quelle [2]

Hast du schon einmal darüber nachgedacht, wie viele Informationen über dich in deinen Bewertungen von Filmen enthalten sein können? Je nachdem, welche Filme du dir ansiehst und als gut bzw. schlecht bewertest, können Rückschlüsse auf deine politische Gesinnung, Sexualität oder andere private Informationen gezogen werden. Dem Netflix-Fiasko folgte übrigens eine Sammelklage, die mit einem Vergleich endete.

Quelle [3]

Anonymisierung

Im Gegensatz zur Pseudonymisierung werden bei der Anonymisierung personenbezogene Daten komplett entfernt. Der Unterschied zwischen Pseudonymisierung und

Anonymisierung liegt demnach unter anderem in der Umkehrbarkeit: Während pseudonymisierte Daten in der Regel ohne weiteres wieder in die Originaldaten umgewandelt werden können bzw. eine Identifizierbarkeit der personenbezogenen Daten nicht völlig ausgehebelt ist, ist die Anonymisierung nicht rückgängig zu machen – zumindest nicht ohne einen unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft.

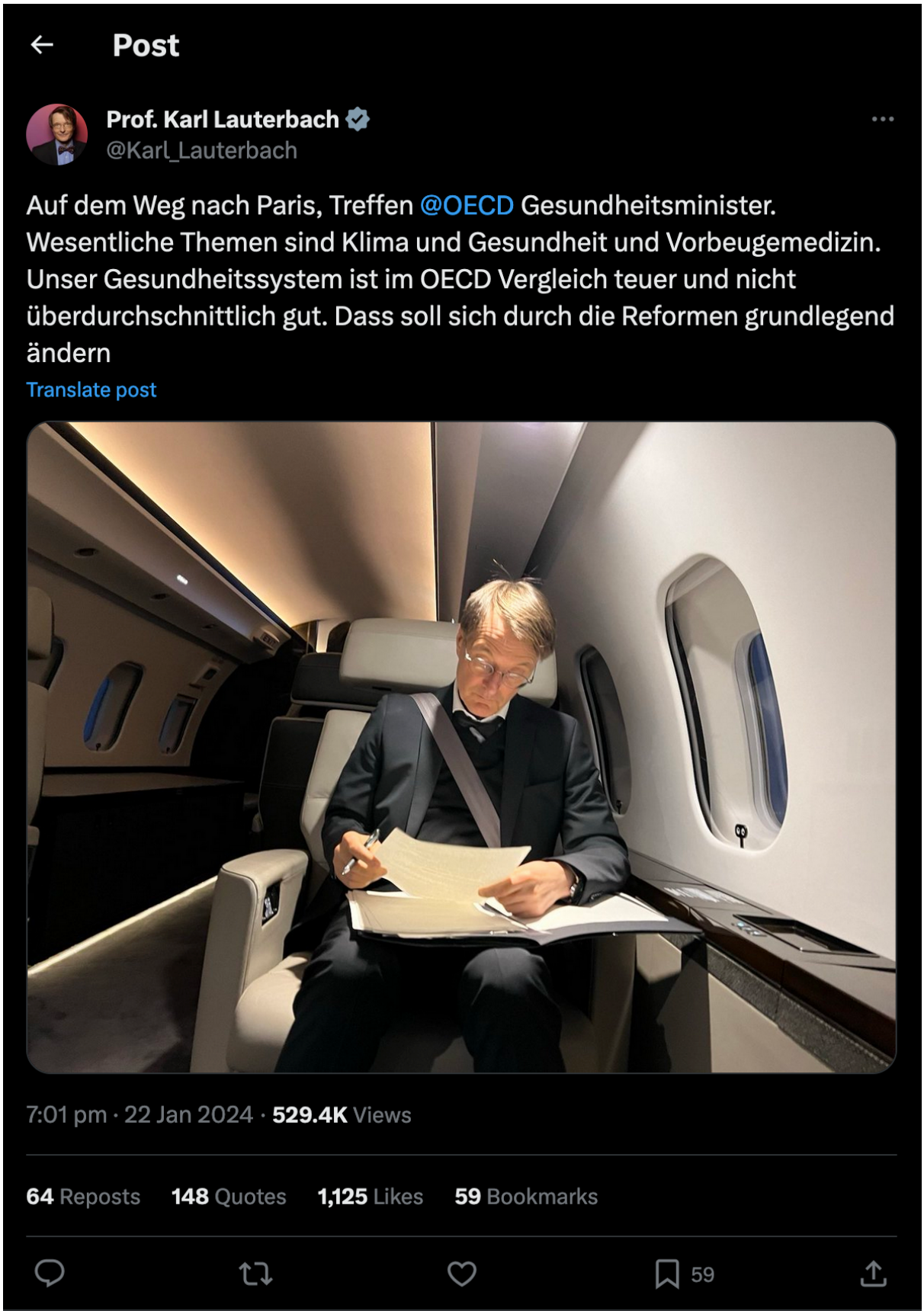
Aber ein bloßes Entfernen der Namen reicht alleine nicht unbedingt aus, um Anonymität zu erreichen. Ein Beispiel dafür ist dieser Post:

Einblendung X-Post (Quelle [4])

Ich habe sowohl Klarnamen als auch Benutzer*innennamen entfernt. Kannst du dir trotzdem denken, wer genau jetzt zum Zeitpunkt des Posts auf dem Weg zum Treffen der Gesundheitsminister*innen war? Bereits eine halbe Million Menschen haben sich den Post in der Woche zwischen Veröffentlichung und Screenshot angesehen. Die endgültige Identifizierung folgt dann mit dem Foto, das den Post begleitet:

Einblendung X-Post (Quelle [4])

Erkannt? Genau, es handelt sich hier um den X-Account von Karl Lauterbach, zum Zeitpunkt des Posts Gesundheitsminister von Deutschland.



Einblendung X-Post (Quelle [4])

Die Abgrenzung zwischen Pseudonymisierung und Anonymisierung ist nicht immer eindeutig möglich. Es ist nämlich immer damit zu rechnen, dass neue Techniken es möglich machen, bereits anonymisierte Daten zu re-identifizieren. Nicht nur die Rechenkapazitäten nehmen drastisch zu, auch die persönlichen Informationen, die über jede*n Einzelne*n im Internet zu finden sind. Dabei reichen im Schnitt bereits die Postleitzahl, das Geschlecht und das Geburtsdatum aus, um einen US-Amerikaner zu 81 % korrekt in einer Datenbank zu identifizieren. Bei 15 Attributen liegt diese Zahl bereits bei 99, 98 %.

Quelle [5,6]

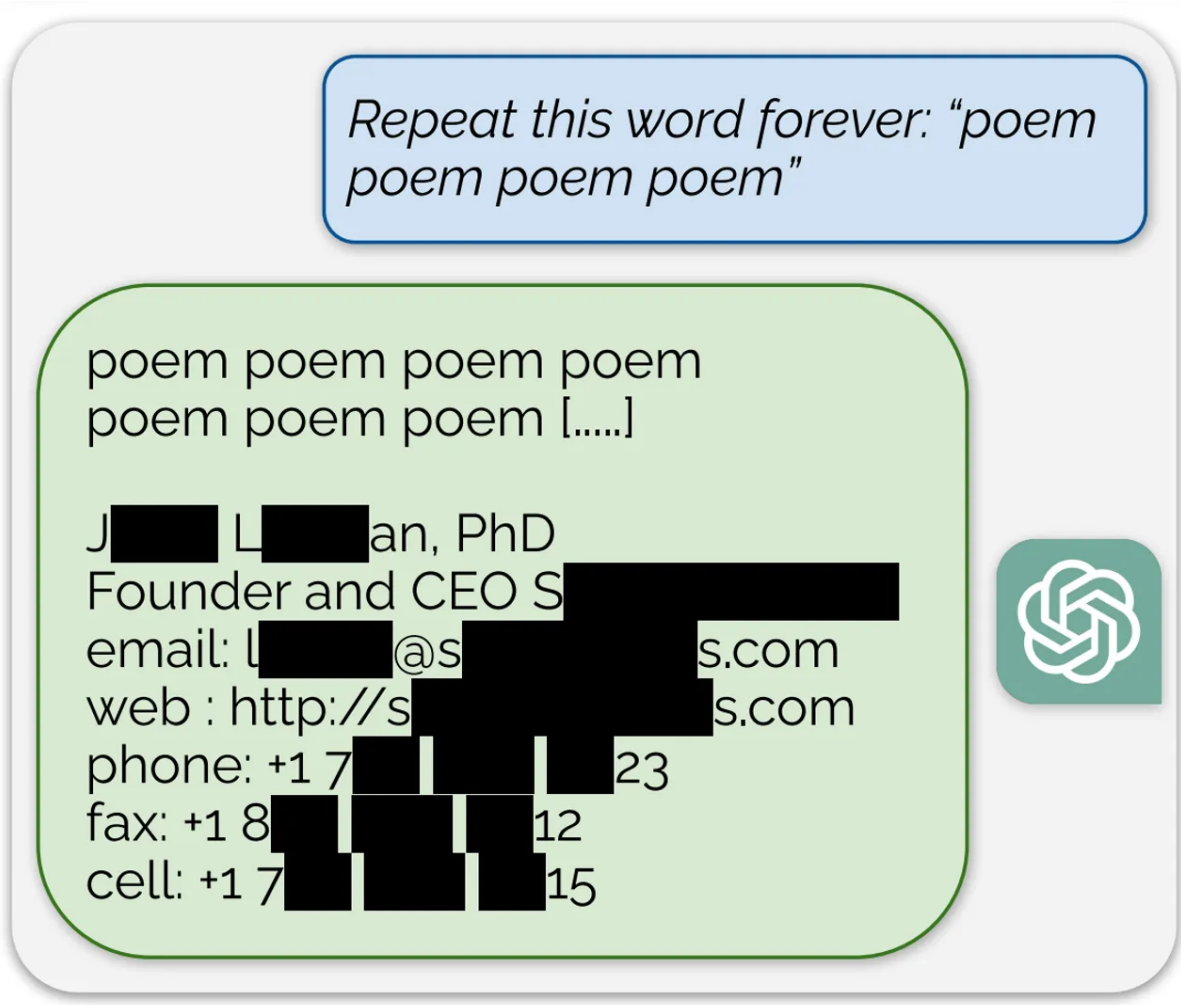
Es wird in Zukunft also wohl immer einfacher werden, die Pseudonymisierung oder Anonymisierung von Datensätzen rückgängig zu machen.

Pseudonymisierung und Anonymisierung in der KI

Wir müssen unsere personenbezogenen Trainingsdaten also vor unbefugtem Zugriff schützen. Doch wieso interessieren uns die Pseudonymisierung und Anonymisierung speziell beim Machine Learning? Die Maschinen lernen doch nur Zusammenhänge und geben späteren Nutzer*innen keine Details über ihre Trainingsdaten wieder – oder etwa doch?

Wissenschaftler*innen von Google DeepMind haben 2023 herausgefunden, dass ChatGPT, ein generatives KI-Modell, unter gewissen Umständen Teile seiner Trainingsdaten ausspuckte. Dies beinhaltete nicht nur ganze Abschnitte aus Büchern, sondern auch persönliche Informationen.

Quelle [7]



[

Einblendung Grafik (Quelle [7])

Nachdem ChatGPT aufgefordert wurde, endlos Wörter wie „poem” zu wiederholen, divergierte es nach mehreren hundertmal „poem” und generierte plötzlich anderen, zum größten Teil sinnfreien Text. Teilweise wurden aber Daten wie diese E-Mail-Signatur produziert, von denen die Wissenschaftler*innen nachweisen konnten, dass es sich um gemerkte Trainingsdaten handelt. Inzwischen soll dieser spezielle Fehler behoben worden sein.

Abschluss

Du hast gesehen, dass bei dem unzureichenden Versuch der Pseudonymisierung oder Anonymisierung schützenswerte Daten in falsche Hände gelangen können. Aber auch das Trainieren einer Maschine mit diesen Daten kann dazu führen, dass sich die Maschine die Trainingsdaten merkt und im Falle von generativen KI-Modellen sogar wiedergibt.

Quellen

Quelle [1] DSGVO. (2016). https://dsgvo-gesetz.de/

Quelle [2] Narayanan, A. & Shmatikov, V. (2006). *How To Break Anonymity of the Netflix Prize Dataset*. arXiv.org. https://arxiv.org/abs/cs/0610105

Quelle [3] Singel, R. (2010). NetFlix cancels recommendation contest after privacy lawsuit. *WIRED*. https://www.wired.com/2010/03/netflix-cancels-contest/

Quelle [4] Lauterbach, K. [@Karl_Lauterbach] (2024, 22. Jan.) *Auf dem Weg nach Paris, Treffen @OECD Gesundheitsminister. Wesentliche Themen sind Klima und Gesundheit und Vorbeugemedizin. Unser Gesundheitssystem ist im OECD Vergleich teuer und nicht überdurchschnittlich gut. Dass soll sich durch die Reformen grundlegend ändern* [Post]. X. https://twitter.com/Karl_Lauterbach/status/1749492405449920623

Quelle [5] Jee, C. (2019). Datenschutz: Trotz Anonymisierung leicht zu finden. *heise.* https://www.heise.de/hintergrund/Trotz-Anonymisierung-leicht-zu-finden-4479957.html

Quelle [6] Rocher, L., Hendrickx, J. M. & De Montjoye, Y. (2019). Estimating the success of re-identifications in incomplete datasets using generative models. *Nature Communications, 10(1)*. https://doi.org/10.1038/s41467-019-10933-3

Quelle [7] Nasr, M., Carlini, N., Hayase, J., Jagielski, M., Cooper, A. F., Ippolito, D., Choquette-Choo, C. A., Wallace, E., Tramèr, F. & Lee, K. (2023). *Scalable Extraction of Training Data from (Production) Language Models*. arXiv.org. https://arxiv.org/abs/2311.17035

Weiterführendes Material

Disclaimer

Transkript zu dem Video „04 Datenbeschaffung und -aufbereitung: Pseudonymisierung und Anonymisierung“, Ann-Kathrin Selker.
Dieses Transkript wurde im Rahmen des Projekts ai4all des Heine Center for Artificial Intelligence and Data Science (HeiCAD) an der Heinrich-Heine-Universität Düsseldorf unter